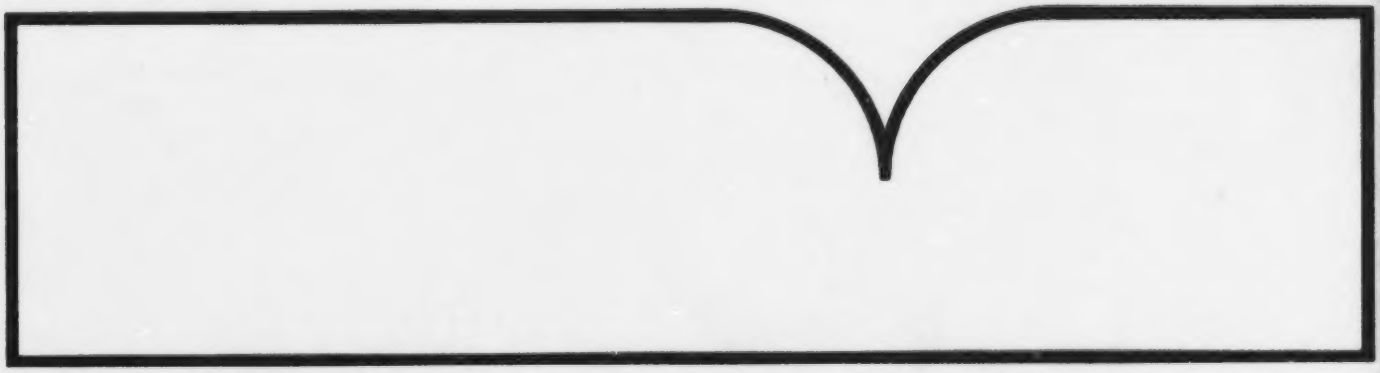
PB80-223704

Procedures for Flow and Auditing of Air Quality Data

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

May 74



U.S. Department of Commerce
National Technical Information Service
NTIS.

BIBLIOGRAPHIC INFORMATION

PB80-223704

Procedures for Flow and Auditing of Air Quality Data.

May 74

PERFORMER: Environmental Protection Agency, Research Triangle Park, NC. Office of Air Quality Planning and Standards.
OAQPS-1.2-013R

The purpose of this guideline is to provide the Regional Offices of EPA with guidance on data auditing techniques that should be followed as part of the procedure for inputting air quality data into the National Aerometric Data Bank. In conjunction with this Guideline, the MDAD is also developing sophisticated data edit, validation and quality control programs which should help smooth the transition between current and planned Regional Office air quality data responsibilities. This report will serve on an interim basis until more explicit and detailed guidance is developed.

KEYWORDS: *Data acquisition, *Air quality data, *National Aerometric Data Bank, *Guidelines.

Available from the National Technical Information Service, Springfield, Va. 22161

PRICE CODE: PC A03/MF A01

PB80-223704

GUIDELINE SERIES

OAQPS NO. 1.2-013R
(Revised 5/74)

PROCEDURES FOR FLOW AND AUDITING
OF AIR QUALITY DATA

U.S. ENVIRONMENTAL PROTECTION AGENCY

Office of Air Quality Planning and Standards

Research Triangle Park, North Carolina

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

TABLE OF CONTENTS

PREFACE		1
1.	Introduction	1
2.	Data Flow Procedures	3
	2.1 Current Data Flow System	3
	2.2 Current Data Editing	7
	2.3 Current Data Validation and Certification	8
	2.4 Current Data Verification	12
	2.5 Future Data Flow System	12
	2.6 Future Data Editing	14
	2.7 Future Data Validation	14
3.	Regional Office Air Quality Data Responsibilities	15
	3.1 Current Areas of Responsibility	16
	3.2 Future Areas of Responsibility	30
4.	Current Techniques for SIP Progress Evaluation	32

LIST OF FIGURES

FIGURE		PAGE
1.	Current Air Quality Data Flow System	4
2.	Future Data Flow	13
3.	Data Anomaly Processing Flow	24
4.	Typical SO_2 Annual Pattern	28
5.	Typical SO_2 Annual Pattern With Constant Baseline Drift	28
6.	Typical SO_2 Annual Pattern With Abrupt Baseline Change	28
7.	Typical SO_2 Annual Pattern With Seasonal Abnormality	28
8.	Influence of Nearby Source on SO_2 Annual Pattern	
9.	Plan Revision Management System	33

PREFACE

The Monitoring and Data Analysis Division of the Office of Air Quality Planning and Standards has prepared this report entitled "Procedures for Flow and Auditing of Air Quality Data" for use by the Regional Offices of the Environmental Protection Agency. The purpose of the report is to provide guidance information on current data auditing techniques that should be followed as part of the procedure for inputting air quality data into the National Aerometric Data Bank. The primary audience for this report is the administrative and management personnel in the Regional Office whose need is limited to a general overview of the system rather than detailed information concerning specific elements. The AEROS (Aerometric and Emissions Reporting System) contact personnel will continue to receive specific detailed information directly from the National Air Data Branch, MDAD. Adherence to the guidance presented in the report will, hopefully, ensure mutually compatible ambient air quality data for all States and Regions and should also facilitate data evaluation and interpretation. Further, any risks involved in policy decisions concerning National Ambient Air Quality Standards should be minimized. This report is intended to update and expand upon the previously issued Interim Guidance Report on "Evaluation of Suspect Air Quality Data."

1. INTRODUCTION

The purpose of this Guideline, the fifth[a] in a series to be issued by the Monitoring and Data Analysis Division (MDAD) of the Office of Air Quality Planning and Standards, is to provide the Regional Offices of EPA with guidance on data auditing techniques that should be followed as part of the procedure for inputting air quality data into the National Aerometric Data Bank.[b] Information and suggestions are presented for both the current and planned computer systems concerning:

- Data Flow
- Data Editing
- Data Validation
- Data Correction Procedures and Certification
- Data Verification
- Statistical Flagging Techniques

In conjunction with this Guideline, the MDAD is also developing sophisticated data edit, validation and quality control programs which should help smooth the transition between current and planned Regional Office air quality data responsibilities.

This report will serve on an interim basis until more explicit and detailed guidance is developed by the Monitoring and Data Analysis Division as a result of the expected interaction with the Regional Offices on air

---

[a] This document supercedes a previously issued interim report entitled "Evaluation of Suspect Air Quality Data" OAQPS # 1.2-006 issued in August 1973.

[b] Information presented in this report is also intended to alert the Regional Offices of their increasing responsibilities with respect to air quality data as a result of the planned upgrading of the EPA/RTP computer system.

quality data handling techniques and procedures. For purposes of definition the following terms are listed as they are used in this report:

Data Check (Data Screen, Screening)

The comparing of a piece of data to a specified entity. The comparison may be manual (visual), or automatic (computerized). The entity may be a code or location (edit) or a value (validation).

Data Auditing

The systematic checking of identifying information and data before or after it resides on the Aerometric and Emissions Reporting System. Includes EDIT, VALIDATION, VERIFICATION, ANOMALY, INVESTIGATION, and CERTIFICATION.

Data Edit (Edit Check)

The comparing of data and its unique identification to a set of specifications concerning format, alphabetic and numeric requirements and coding requirements, etc., either manually or automatically.

Data Validation (Validation Screen)

The comparing of data values to a set of predetermined criteria concerning minimum and maximum limits, deviation from average values, percent change overtime, etc., either manually or automatically.

Data Anomaly (Anomalous Data)

Any data or data summary about which some problem exists or about which there arises a question as to its integrity of

Data Flag (Flagging)

Calling attention to and uniquely identifying data for futher action, the flagging maybe done manually or automatically.

information. Anomalous data may be identified (flagged by a report) either manually or automatically by edit checks, validation or any other flagging technique.

Data Verification

The total process involved in determining the existence of data which, while not on NADB, has been indicated as existing by knowledgeable sources.

Data Certification

The process by which data currently residing on NADB is determined to be correct and complete or is recoded by individuals sufficiently knowledgeable to have background authority and data to represent the source.

2. DATA FLOW PROCEDURES

This Section presents the current procedures for processing air quality data. These procedures include, as required, data editing, validation, verification, certification and flagging techniques for SIP progress evaluation.

2.1 Current Data Flow System

The general flow of air quality data from the States through the Regional Offices to the National Aerometric Data Bank is presented in Figure 1. The steps in the system are as follows:

a. The State agency submits air quality data to the appropriate EPA Regional Office as part of the State Implementation Plan reporting procedures. These reports which are forwarded on a quarterly basis contain the air quality

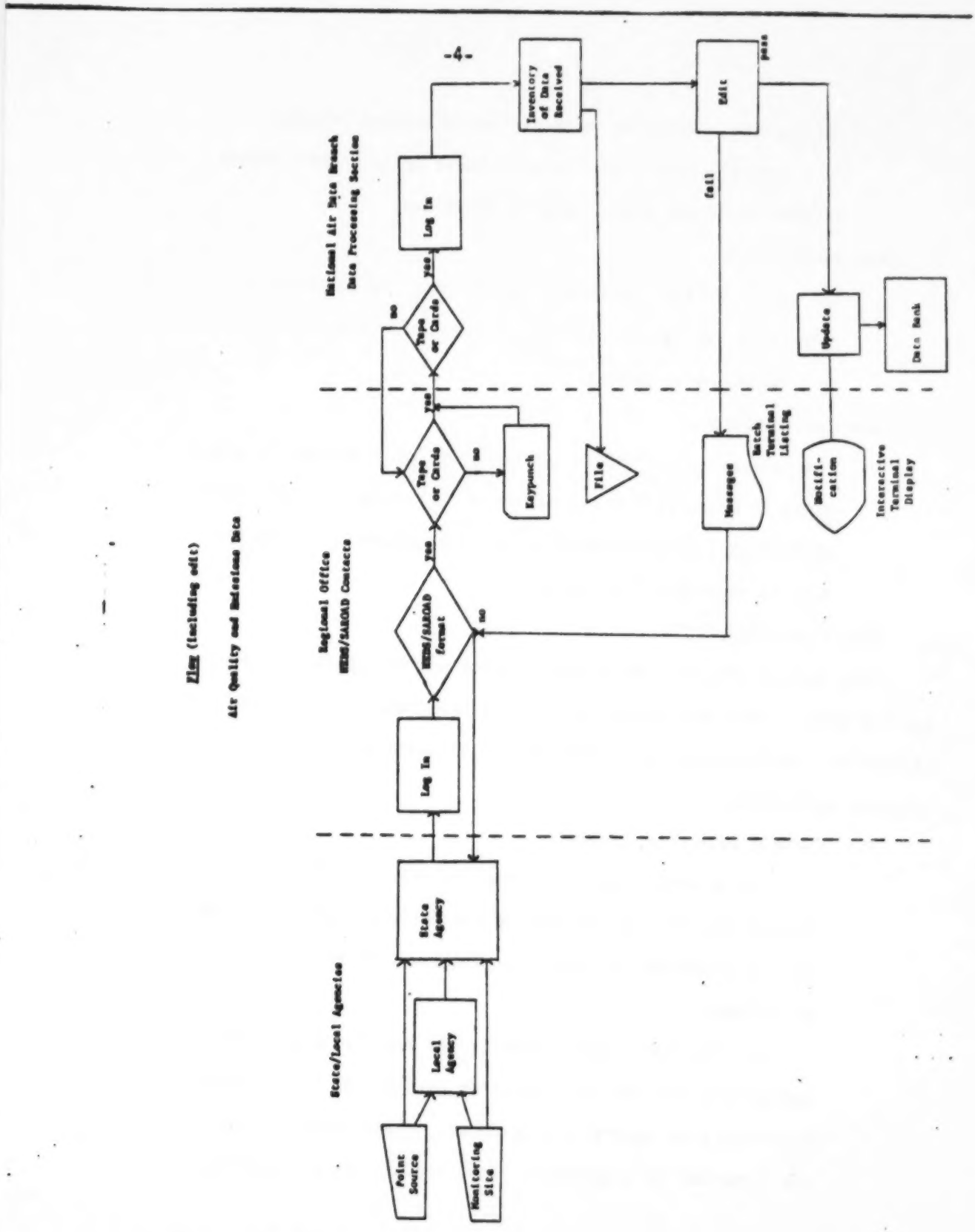


Figure 1 - Current Air Quality Data Flow System

data and new site descriptions for the State's air monitoring stations. The data may be sent in more frequently than quarterly if desired, but must be submitted to the Regional Office in SAROAD format on either coding forms, punched cards, or magnetic tape. Data for all operational stations as described in the SIP's, beginning with that used in plan preparation, must be submitted. It is strongly encouraged that all reliable data obtained by the State which satisfies the criteria established for monitoring network adequacy be submitted.

b. The NEDS/SAROAD contact in the Regional Office arranges for keypunching of forms if necessary and then mails the data to the MDAD's National Air Data Branch in card or tape form.

c. Air Quality data submitted to the National Air Data Branch should have the following characteristics:

i. Data must be coded in SAROAD format.

ii. Data values less than the monitoring minimum detectable sensitivity should be reported as a "zero" value. A value equal to half the minimum detectable sensitivity will be substituted when calculating summary statistics for continuous data.

iii. It is desirable that the data be representative of a consecutive three-month period for which at least 75 percent of the data values are valid. A non-detectable measurement, i.e., a value below the minimum detectable sensitivity (Limits of Detection),

is considered valid. Summary statistics are not automatically machine computed if greater than 50 percent of the valid measurements are below the minimum detectable concentration. However, if the criteria are not met, the data should still be submitted particularly for evaluation of maximum value standards. For noncontinuous 24-hour data there should be at least five data points in the quarter, with at least two months being reported and a minimum of two data values in the month with the least number of data value reported.

iv. Data must represent an interval of one-hour or greater -- shorter interval data must be averaged over an hour.

v. Data must be representative of the conditions of the site for the period of time specified; modification of the environment in which the site is located must be reported to the MDAD by the State and/or the Regional Office.

d. Data are processed using the SAROAD edit program and the error messages generated are provided to the AEROS contact.

e. Investigation and correction of potential errors is accomplished by the Regional Office in conjunction with the States using procedures described later in this document. Corrected data are submitted to the National Air Data Bank for file updating.

2.2 Current Data Editing

The incoming air quality data, in SAROAD format, is subjected to various checks by the National Aerometric Data Bank's computer programs. The data will fail to pass the edit programs for the following reasons:

a. No existing site description. Before any data are accepted, the site file must contain the information from the site identification form. The program checks the 12-digit site code on the data and if no corresponding record is available in the site file, the data are rejected. Therefore, the site identification must be entered before data from a new site can be accepted.

b. No existing description of sampling or analytical method. The program automatically rejects data if a record of the method used to generate the data is not available.

c. No match on the pollutant-method-interval-unit combinations for these codes. Anything else will be rejected. For example, there is no monthly interval suspended particulate data using a hi-vol sampler and gravimetric analysis.

d. Any data field other than "Agency" or "Interval" which has been coded in alphabetic rather than numeric characters.

e. Data on the wrong form, such as trying to send 24-hour data on the hourly data form.

f. Incorrect start hour. For hourly data the start hour must be 00 or 12. For two-hour data through twelve-hour data

legitimate values are given on page 36 of the SAROAD Users Manual.[1] For twenty-four hour or greater data, legitimate values are from 00 to 23. Anything else is automatically rejected.

g. Data incorrect. Data are checked for meaningful days. Examples of meaningless days are February 30 or April 31. Some data had to be rejected because the year was designated as 1977. Eventually, the capability to flag data which have a date other than the current quarter will be added. However, this capability will be delayed until all back data are incorporated in the system.

h. Imbedded non-numeric characters in values. There is a four digit field for the value. For example, values which have blanks between digits, such as two zeros, a blank, and an eight instead of three zeros and an eight would be rejected.

i. Decimal place indicator not between 0 and 5. The data which are currently being generated all have fewer than five decimal places.

2.3 Current Data Validation and Certification

Currently, the manual procedure used by the MDAD in the identification of potentially anomalous data values depends, to a large extent, on chance discovery by someone scanning a computer printout of either raw data or summary statistics. Automatic procedures have not yet been developed for computer applications.

This process of detecting questionable data values will be supplanted when the data system is transferred to the Univac computer in August, 1974. Potentially anomalous values will be objectively identified as a step in the addition of all new data to the file. Both parametric and non-parametric tests could be applied to the incoming data and a listing printed of all values that meet one or another of the test criteria for flagging. Examples of such tests are given below.

Non-parametric tests

- Values that are larger than the arithmetic mean of the data by some preassigned factor (such as 2).
- Values that are some factor, say 1.5 times larger than the estimated assigned 99th percentile of the data.
- Hourly values that differ from adjacent values by more than some preassigned ratio, suggesting some abrupt change in baseline or a transient interference.
- Chebyschev type tests, wherein values that are more than four standard deviations away from the mean are to be considered suspect.

Parametric tests

Efficient use of these tests depends on knowledge of the frequency distribution of the quantity being measured. Example of such tests are presented below. (The sensitivity of these tests can be determined analytically from the frequency distribution.)

- Detection of any values that are larger by some factor (e.g., 1.5) than the expected value of the assigned 99th percentile of the distribution under question.
- The finding that the average of $K \geq 5$ successive values falls outside the $(\mu \pm \frac{3\sigma}{\sqrt{K}})$ limit, where μ and σ^2 are the mean and the variance of the distribution under question.

Note: The difference between the non-parametric test and the parametric test is that in the former, the assigned percentile is estimated from the data, whereas in the latter it is theoretically obtained.

Validation of the pollutant measurements involves technical judgment about what constitutes questionable data, and is expected to be applied systematically in the form of a set of criteria defining, for each pollutant, what constitutes an unusual or anomalous value or an abnormal fluctuation. Excursions outside of expected bounds should be flagged or tabulated but cannot be automatically rejected or deleted. They must be brought to the attention of the contributing agency for correction.

Definitions of what constitute unusual values or abnormal fluctuations are required for each pollutant. These criteria should be defined by people familiar with the characteristic behavior of the pollutants and the instruments used to measure them. Realistically, these criteria for identifying questionable values should be open to revision. Once developed, these criteria can be readily incorporated as a standard element in the data bank's editing and/or validation procedures.

Certification by States is accomplished by using available SAROAD output to determine the accuracy and completeness of all submitted data. Particular emphasis should be placed on the following:

a. Site identification information

b. Methods of collection and analysis

c. Integrity of the actual data

All three items must be coded and represented on the data bank as accurately as possible to insure the proper interpretation and evaluation of the data.

Certification may be triggered by either of two mechanisms: First, any time there are FDIT or VALIDATION reports flagging either incorrect data or data of a questionable nature, implicit certification is required. This means that the data must be corrected and resubmitted, if necessary; otherwise, for data which has been flagged as being possibly invalid, no action is necessary if the data is correct as it was submitted.

The second trigger for certification may be dependent upon time or the number of anomalies being reported for a specific subset of data. It may be determined that an agency should inspect a set of data to certify it as being correct and complete. In this situation, and it will always be identified as such, the appropriate agency must make any corrections necessary to the data and must always respond in writing that the data are correct as they stand or that the corrections which have been attached will solve the problem.

2.4 Current Data Verification

Currently the entire procedure of data verification is being handled through contractual resources. This involves the use of reference publications to determine the probable existence of additional air quality data. Once NADB is aware of this data the necessary steps are taken with the appropriate agency to coordinate the submission of the data to the National Aerometric Data Bank.

2.5 Future Data Flow System

As previously mentioned, it is expected that the Regional Offices will assume more responsibility with respect to the validation of air quality data. This will be accomplished by their taking a central role in the screening of air quality data before it is inputted into the National Aerometric Data Bank. The screening will involve not only editing the coding format but also the validation of the measurements.

During the transition period of shifting more responsibility to the Regional Offices, it is anticipated, at least initially, that the MDAD will do minimal revalidation of the data. Also, the flagging technique for measuring SIP progress will still be employed and the National Air Data Branch will assume the ultimate responsibility of entering the "correct" SAROAD data into the National Aerometric Data Bank (Figure 2).

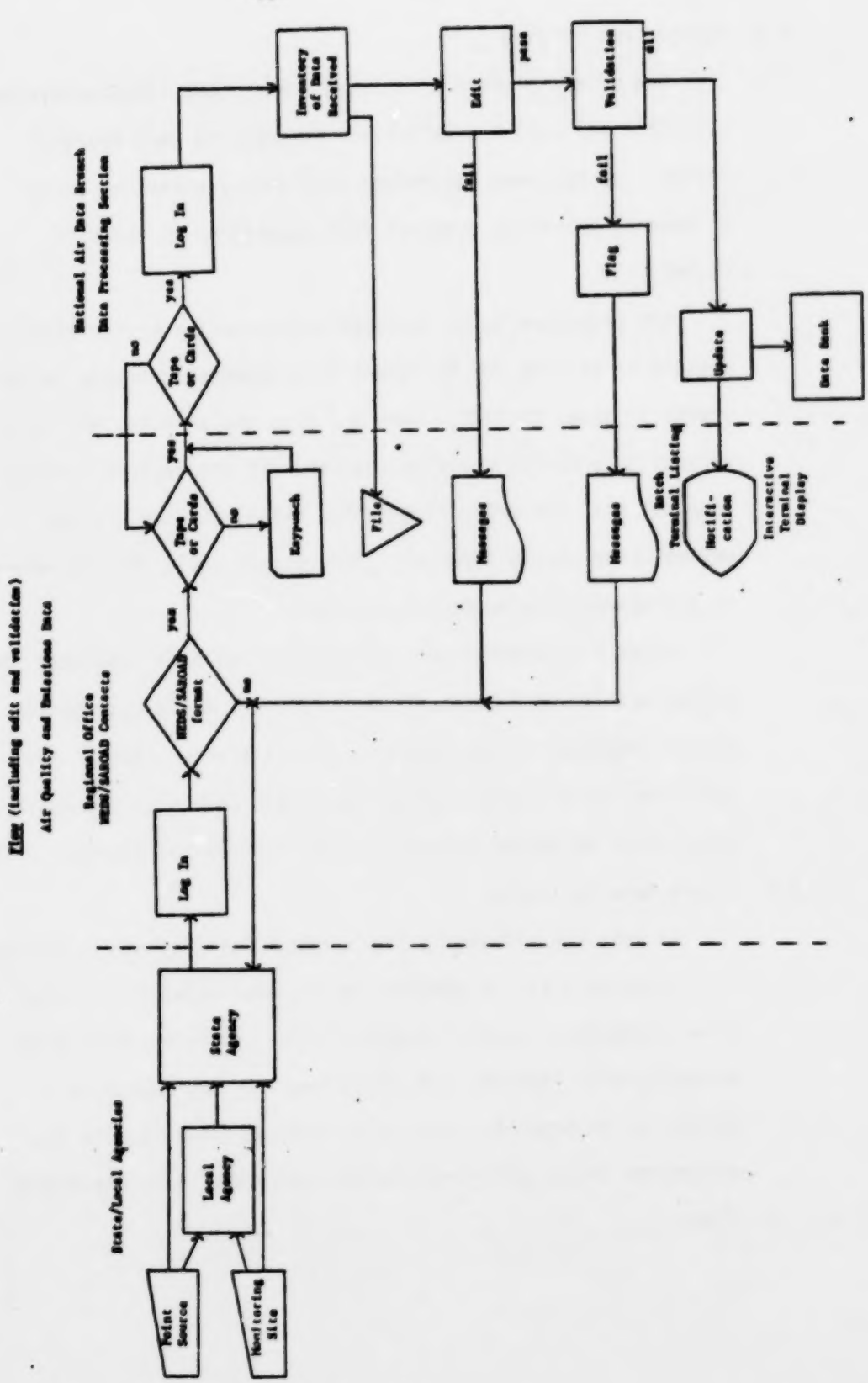


Figure 2 - Future Data Flow

2.6 Future Data Editing

One of the highest priorities within MDAD concerns making available all Edit and Validation programs to each Regional Office. It has been determined that this can best be accomplished by providing terminal edit capability on the RTCC-UNIVAC 1110.

The procedure to be followed would involve either transmitting or mailing the AQ report in a computer readable medium (cards or tape) to RTCC. Once the data has arrived, the edit/validation programs could be executed via the Regional Office terminal with the error diagnostics being returned via the medium speed remote terminal. This output could then be returned to the appropriate agency as required.

After a successful edit of the data has been completed the culled data would be identified to NADB who would concatenate several Regional Office data sets into a single update. Any additional errors generated by the actual update (i.e., duplicate data) would be routed directly to the appropriate Regional Office.

2.7 Future Data Validation

As data are audited by the Terminal Edit/Validation program it is planned that, in addition to the edit rejection listing being produced, a special report will be generated which automatically will identify data which seem for one reason or another to be invalid. This data although identified in the validation report will nevertheless be updated onto the SAROAD files.

Due to storage constraints there are no plans for these data to be further "flagged" while stored. It is imperative that the data be checked immediately to determine its validity by the submitting agency. If the data are confirmed to be correct no further action is necessary. If, however, the data are incorrect then the agency must immediately code the necessary changes and/or deletions and submit these to the appropriate Regional Office.

In addition to the types of validation tests already discussed the following list illustrates the computerized hourly validation checks under consideration:

CO	100 ppm
SO_2	2 ppm
Ozone (Total Oxidant)	.7 ppm
Total Hydrocarbons	10 ppm
Non-methane Hydrocarbons	5 ppm
NO_2	2 ppm
NO	3 ppm
NO_x	5 ppm
Total Suspended Particulate	2000 g/m^3

3. REGIONAL OFFICE AIR QUALITY DATA RESPONSIBILITIES

This Section presents recommendations and suggestions as to those methods and techniques which the Regional Offices can employ to validate air quality data. The Monitoring and Data Analysis Division recognizes that some of the areas of responsibility are beyond the capability of some of the Regional Offices at this time. In these cases, the MDAD will

provide technical and other assistance on an as needed basis in order that the current and planned data flow system operate in the most efficient and effective manner possible.

3.1 Current Areas of Responsibility

At this time, there are various tasks which the Regional Offices perform in the validation of air quality data. These include the following:

a. Preliminary Data Inspection

The Regional Office can make a visual screening of the SAROAD sheets before forwarding the data to MDAD. Ensuring that the site identification and descriptions, pollutant, sampling and analytical method, interval, units and decimal point locations are properly filled in on both the 24-hour and hourly SAROAD coding form will greatly reduce the edit and resulting correspondence between MDAD and the Regional Offices. If a particular agency shows a history of carelessness in correctly filling out their SAROAD sheets, the Regional Office may want to check these sheets for their "correctness" as discussed in Section 2 rather than just for their completeness.

If the data submitted to the Regional Office from the States are in the form of punched cards, the Regional Office can visually inspect the batch to make sure that pertinent columns are punched and aligned correctly. The Regional Office may find it desirable to actually print out or list the data from selected agencies before forwarding the cards

to MDAD. If the data are sent on magnetic tape, there is little the Regional Office can do, at present, but forward it on.

b. Interrogate Data Bank, Data Requests and Manual Examination

Some existing SAROAD outputs are available which the Regional Office may find helpful in evaluating their air quality data. The Regional Office can request output from the data bank and get quarterly and yearly frequency distribution lists for each sampling station. The output includes the site description at the top of each page and a frequency distribution for each pollutant, year or quarter-year. The number of observations, minimum, maximum, and the percentile values are listed for each pollutant-quarter-year. The arithmetic mean, geometric mean, and geometric standard deviation are given only for those pollutant-quarter-years which meet National Aerometric Data Bank criteria.

The frequency distributions are available on a national, EPA regional and State basis. Other options include the ability to request the distribution for limited numbers of pollutants, years or quarters.

These and other outputs and remote batch and interactive access methods are more fully defined and discussed in the SAROAD Terminal Users Manual,[2] and the Regional Office NEDS/SAROAD contact should be contacted for additional information.

The Regional Office will, in the future, be able to make comparisons between measured air quality data and that which they, and/or the State and local agencies, intuitively feel is reasonable for that geographical area, station and pollutant.

c. Check Anomalous Data

Anomalous or questionable data values may arise from the data flow system as a result of the following procedures: edit checks, validation screen and the application of the flagging technique. The Regional Office has the responsibility of either accepting, rejecting or modifying the data value or average in question. In this regard, the Regional Office has the option of requesting that the originating agency determine the validity of the data or provide certain information and documentation so that they may make the final determination.

The procedure used to check out any specific data value prior to the initiation of an anomaly request to NADB could depend on: the Regional Office's assessment of the originating agency in terms of its capability, quality control program, and previous performance. MDAD suggests that the following sequence of steps be followed in order to check out anomalous data values or composite averages. In all cases, it should be recognized that any agency which alters, manipulates or transcribes a data value in any way is potentially capable of introducing an

error. When a data value is identified as being questionable, the responsible agency must determine whether or not the data value maintained its integrity throughout the agency's data acquisition and processing system.

The data should be traced through the SAROAD system, the Regional Offices, State agency and/or local agency to its original recording, whether it be a value from a computer readout, paper tape printer, strip chart, or a report from the chemist in the laboratory. The types of errors usually found in the internal check are: typing, key punching, tabulating and transposition, mathematical (such as addition, multiplication and transcribing). Further discussion of these errors and methods to reduce their frequency may be found in already published guideline documents.[3,4,5]

If no errors have been identified in the internal check, at all agency levels, the verification and evaluation process should continue down two similar but separate paths. Which path is chosen depends on whether the data in question is a single value or a composite average.

1. Evaluating Specific Air Quality Data Values

   · Instrument Calibration, Specifications and Operations

   The operation and calibration of continuous instruments is of the utmost importance in the production of valid air quality data. The instrument calibration should be reviewed for the time period in question, both before and after the suspect data

point. It should be determined if the instrument was operating within pre-determined performance specifications such as drift, operating temperature fluctuations, unattended operational periods, etc. These performance specifications for automatic monitors are defined and published in the Federal Register[6] and summarized in various guideline documents.[3,4] These specifications are likely, however, to be superceded by those published in the October 12, 1973, issue of the Federal Register on proposed Equivalency Regulations. Guidelines on air quality control practices and error tracing techniques are also available.[5]

Before and After Readings

If the instrument generating the data was found to be "in control," the values immediately before and after should be determined. Comparisons between the percent and/or gross deviations could be made. Ideally, this difference in concentration should be determined through a statistical analysis of historical data. For example, it may be determined that a difference of 0.05 ppm in SO_2 concentration for successive hourly averages occurs very rarely (less than one percent of the time). The criteria for what constitutes an excessive change may also be linked to the time of day.

For example, an hourly change of CO of 10 ppm between 6 AM and 7 AM may be common but would be suspect if it occurred between 2 AM and 3 AM.[3,5]

- Other Instruments at the Same Location

  Observing the behavior of other instruments at the same location would give the evaluator a qualitative insight into the possible reasons for the anomalous reading. If all of the instruments showed a general increase, meteorological factors might be considered while a dramatic deviation over the same short period of time may indicate an electrical problem or an air conditioning malfunction. On the other hand, if the other instruments behaved normally, a temporary influence of a single pollutant or single pollutant source may be suspected.

- Similar Instruments at Adjacent Locations

  Comparing the behavior of other instruments in the vicinity which monitors the same pollutant could further elucidate the situation. For example, if the adjacent instruments (upwind and downwind) exhibited the same general trend, an area problem in which the maximum effect was over the station of interest, would be indicated. However, if the adjacent stations seemed to peak either before or after the time the suspect value was recorded, the station may have been under the influence of plume fumigation

which wandered according to wind direction influences. Micrometeorological influences should not be overlooked either. The station may be under the influence of subsidence effects from the urban heat island or upslope-downslope influences.[7,8]

- Meteorological Conditions

  No attempt to explain an anomalous air quality data point would be complete without a consideration of the meteorological conditions present at the time of the reading. A passing front and strong inversion, extended calms or strong winds are conditions which have a great impact on air quality.[7,8] Influences of precipitation, temperature and season could be included to interpret the reasonableness of the data as well.

- Time-Series Check

  Investigating a time series plot of the data might reveal a repetitious pattern during similar time periods. An extreme excursion might thus be explained. For example, the instrument may be extremely temperature sensitive and may be under the influence of the sun shining between buildings from 2 PM to 4 PM each afternoon. Similarly, for example, every Thursday may be delivery day for an adjacent supermarket where the delivery trucks spent the bulk of the day idling in the vicinity of the sampler probe.

- Physical Site Location

From time-to-time local air quality influences may change and adversely affect a given air monitoring station's representativeness. Examples of this might be an adjacent apartment house or supermarket changing from garbage haul-away to an incinerator. Urban renewal may also render the location temporarily unrepresentative. It may be beneficial for each agency or Regional Office to maintain a map and photograph of each site showing influencing site characteristics. These could be updated on a periodic basis. The site location, sampling probe material and configuration should also be within the bounds of those specified in published guidelines.[3] Figure 4 presents a stepwise review and guide to the verification of specific data values. It should provide the Regional Offices with an overall picture of the suggested processing of State and local air quality data.

ii. Evaluating Annual Air Quality Averages

- Summary Statistics

If no calculation or recording errors have been found, those summary statistics which describe the average should be checked. These may include both geometric and arithmetic means, standard deviations, and the frequency distribution in percentiles. Both the

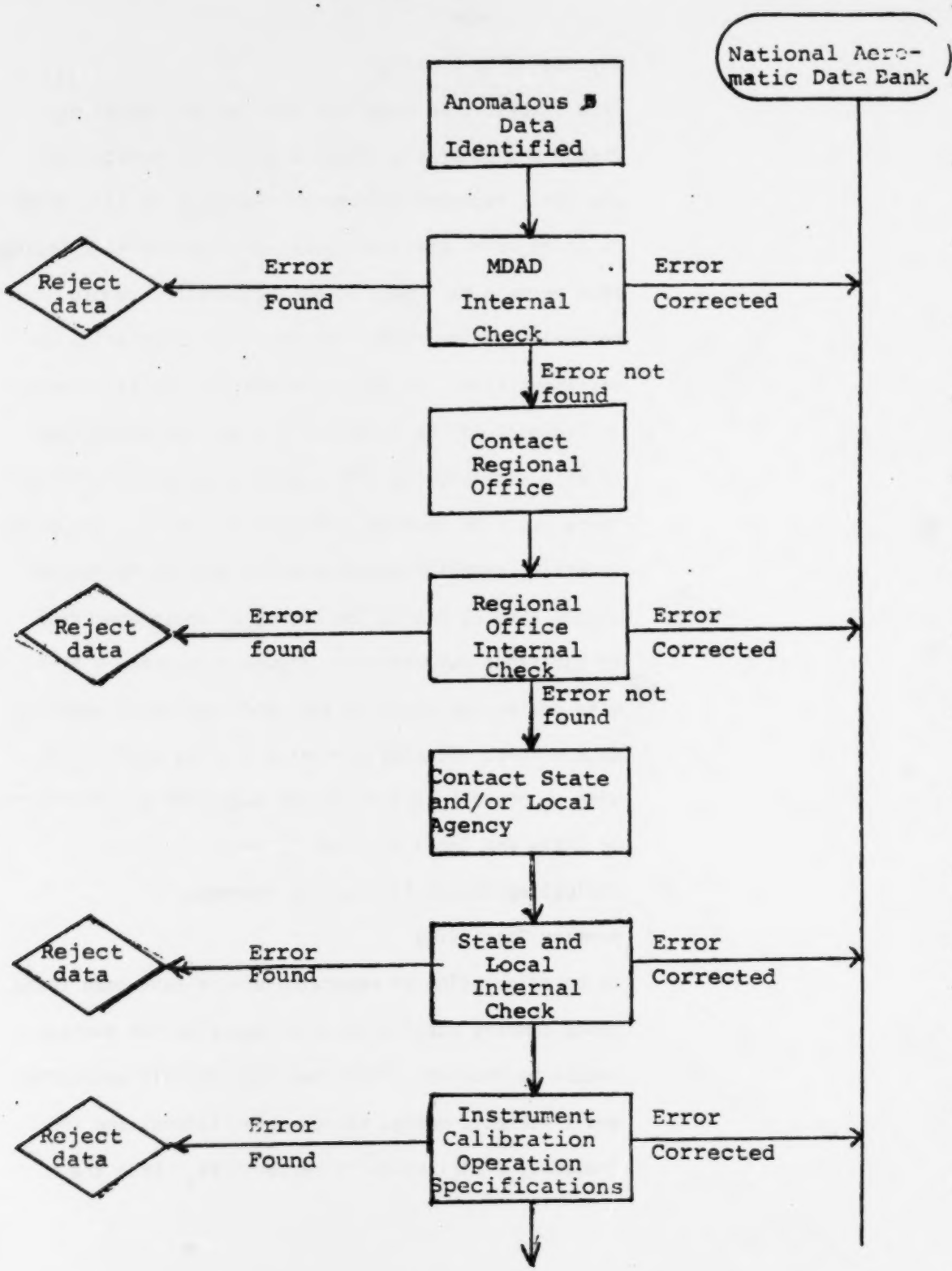
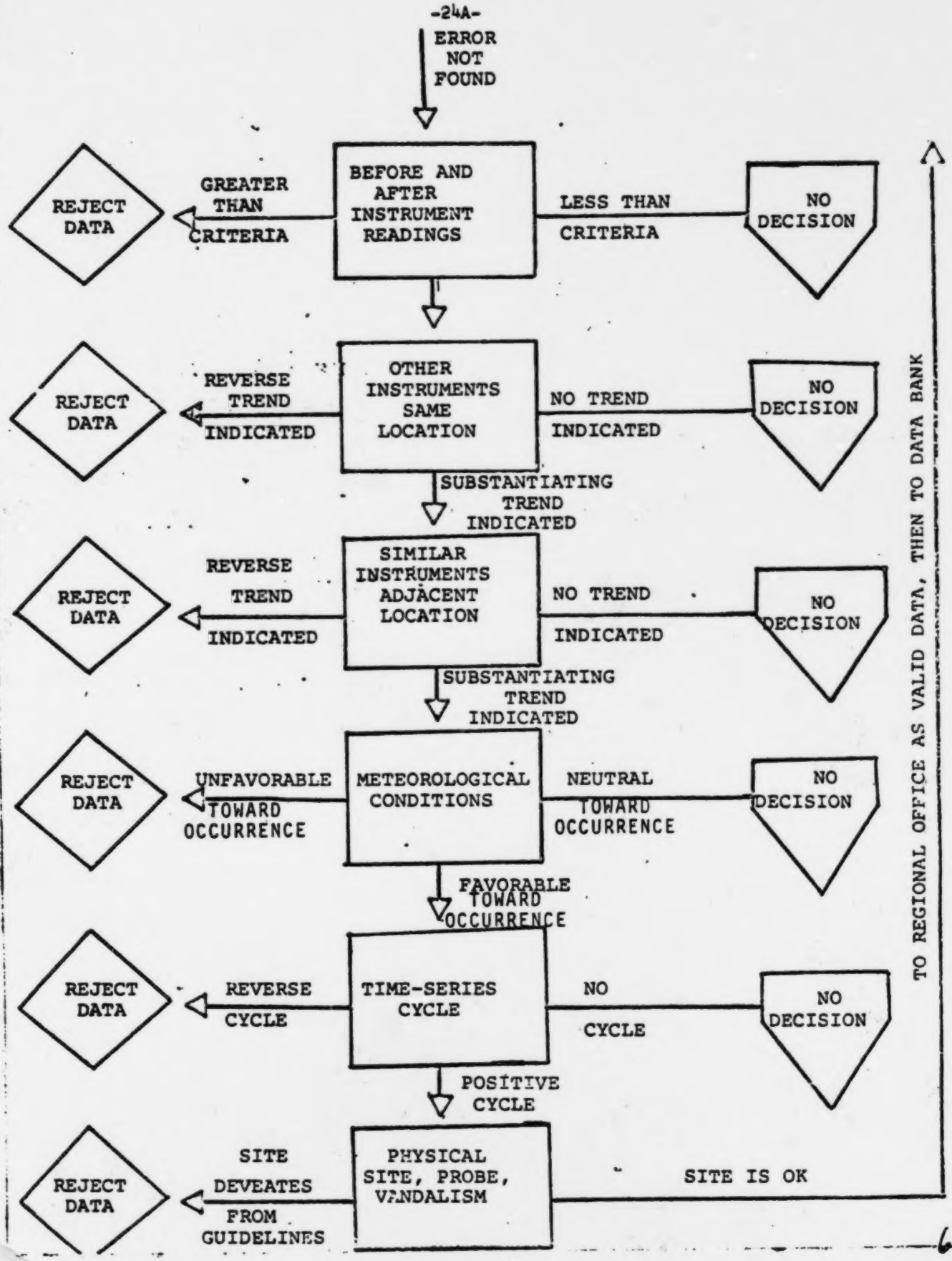


FIGURE 4. DATA VERIFICATION FLOW CHART FOR SPECIFIC DATA VALUES

66


ERROR NOT FOUND
BEFORE AND AFTER INSTRUMENT READINGS
GREATER THAN CRITERIA
REJECT DATA
LESS THAN CRITERIA
NO DECISION
OTHER INSTRUMENTS SAME LOCATION
REVERSE TREND INDICATED
REJECT DATA
NO TREND INDICATED
NO DECISION
SUBSTANTIATING TREND INDICATED
SIMILAR INSTRUMENTS ADJACENT LOCATION
REVERSE TREND INDICATED
REJECT DATA
NO TREND INDICATED
NO DECISION
SUBSTANTIATING TREND INDICATED
METEOROLOGICAL CONDITIONS
UNFAVORABLE TOWARD OCCURRENCE
REJECT DATA
NEUTRAL TOWARD OCCURRENCE
NO DECISION
FAVORABLE TOWARD OCCURRENCE
TIME-SERIES CYCLE
REVERSE CYCLE
REJECT DATA
NO CYCLE
NO DECISION
POSITIVE CYCLE
PHYSICAL SITE, PROBE, VANDALISM
SITE DEVEATES FROM GUIDELINES
REJECT DATA
SITE IS OK
TO REGIONAL OFFICE AS VALID DATA, THEN TO DATA BANK

standard deviations and the magnitude of the difference between the geometric and the arithmetic means are more sensitive to a few extremely high values than to many moderately high levels. Inspection of the values corresponding to the higher percentiles would also show the influence of abnormally high values. On the average, standard deviations do not generally change much from year-to-year.

- List Individual Values

  If the summary statistics indicate that the mean was heavily influenced by a few high values, or in the absence of summary statistics, the individual data values which comprised the average should be listed. From inspection of this list, it can be determined if the average was influenced by a relatively few large values or whether the bulk of the data appears to be consistently high. If the former appears to be the situation, each individual data value should be treated according to the guidelines for specific air quality data points presented above. In the latter case, proceed to the next step in the verification of annual averages.

- Physical Site Inspection

  The physical site location should be evaluated in terms of its representativeness of the pollutant of interest,

the averaging time of interest, and the pollutant receptor. The operation of the site should be evaluated in terms of sampling methodology, maintenance procedures, calibration procedures and quality control practices. The actual sampling probe and manifold material, configuration and placement should also be evaluated. Guidelines describing in detail these aspects of air quality monitoring have been published.[3,4,5]

- Plot Data

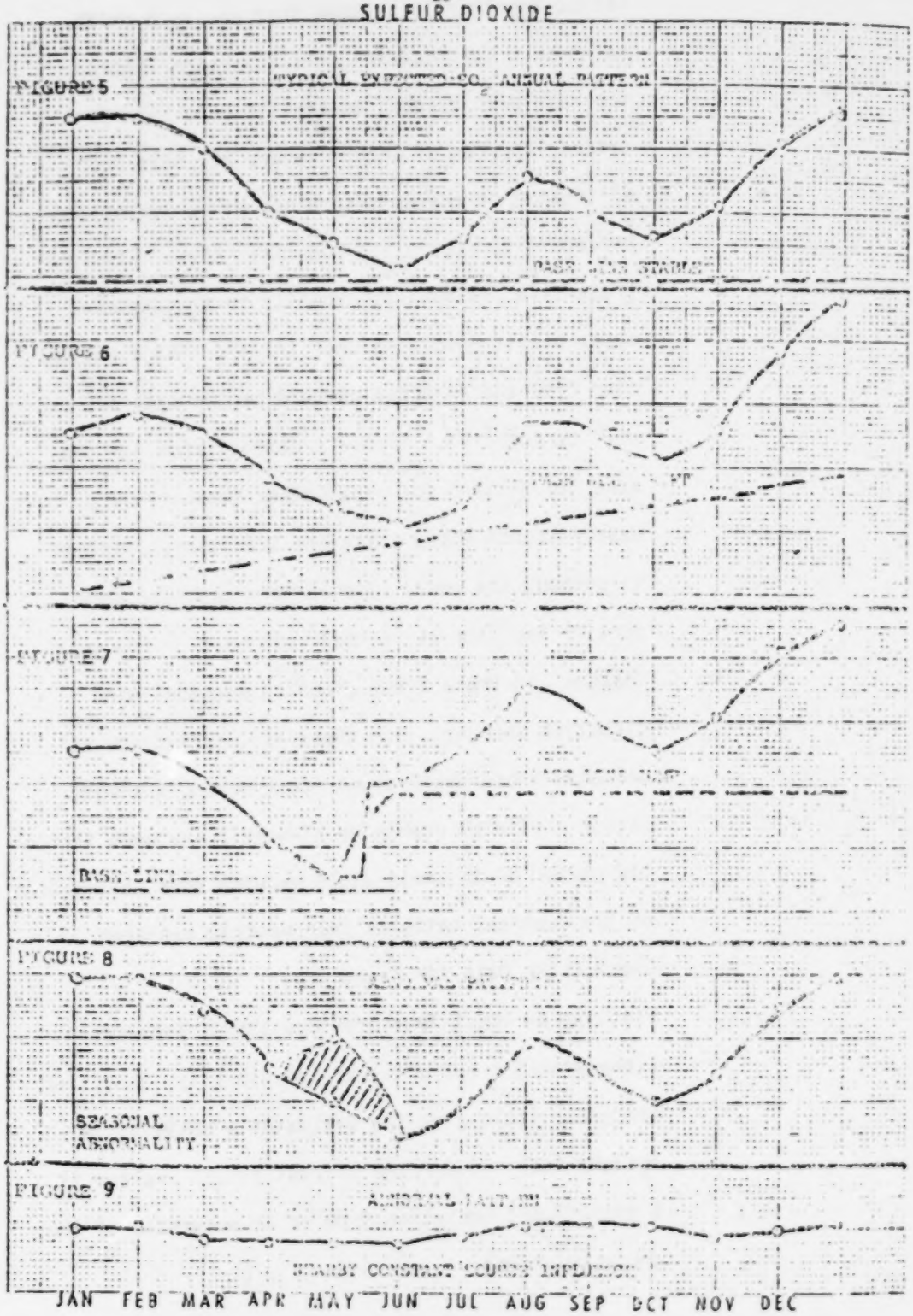
Comparing a visual plot of the current data to that of prior years on a typical annual pattern could further pinpoint reasons to accept or reject the annual average in question. Note that, however, some year-to-year variation is expected. Figure 4 presents a typical SO_2 annual pattern based on expected monthly averages (exaggerated for purposes of illustration). Figure 5 also shows this same pattern with a constantly increasing baseline drift. A pattern of this type suggests a continuing long-term failure (change) in a component of the instrument, deterioration in the supplies being used or a subtle change in the environment. Figure 6 presents the typical pattern with an abrupt dislocation of the base line. This may be indicative of a change in struments, methods of analysis, procedures used or personnel. It should

not be arbitrarily assumed that any such shift is wrong. For instance, the analytical method may have been changed to the standard reference method, sources of interferences may have been eliminated or the operators may be following the procedure correctly for the first time. Figure 7 presents a seasonal abnormality in the expected pattern. It should be kept in mind that a deviation from the expected pattern can be negative as well as positive. Figure 8 demonstrates how the expected pattern can be smoothed (masked) by a nearby source whose emissions are fairly constant throughout the year. The pattern may also show part of the year as "normal" and part of the year "masked" if there are pronounced seasonal wind direction changes. For those pollutants such as oxidants whose peak values occur during a single season a plot of weekly or bi-weekly averages through the period of interest would provide more information on the cyclical patterns than monthly averages.

Check Prior Data for Trend

Plotting at least four previous annual averages along with the current year and visually inspecting the graph could give the evaluator a qualitative insight into whether the current annual average is a significant deviation from or an extension of the projected trend.

SULFUR DIOXIDE



Reproduced from best available copy.

72

- Compare With Surrounding Stations

  If there are enough surrounding sites to develop air quality isopleths of the area, the evaluator could see how the annual average in question fits in with the overall picture. For instance, if the point in question was midway between the isopleth lines representing 80 and 60, but the recorded value was 50% greater than expected, i.e., 105, an abnormality may be expected.

  This comparative technique may also be used in areas where there are not enough sites to directly plot air quality isopleths but where a predictive air quality model has been developed and verified with a limited number of actual data values. In these cases, for example, deviations of $\pm$ 100% could be suspect.

- Meteorology

  The annual average should be interpreted in conjunction with meteorological conditions for the year in question. For example, if the winter of the year in question were the coldest in 50 years or the overall degree days were 50% above the 20-year norm, an increased SO_2 average would be expected. Suspended particulate values can be greatly affected by wind direction and a disproportionate wind rose (atypical for the area) could help explain unusual values. Comparing the appropriate meteorological parameters

such as rainfall, wind speed, number and length of inversion, temperature and degree days to their long-term averages, i.e., 20- or 50-year norms, before attempting to change implementation plans is suggested.

d. Data Bank Add/Correct/Delete Procedures.

As Regional Office interaction with the SAROAD data bank increases, there will be an increasing need to become proficient with the procedures used to update the bank with new data, correct existing data and delete data which are incorporated in the data bank but have been found to be in error. There are then three types of transactions which can be processed by the SAROAD data bank: add, correct, and delete. In each case data in SAROAD format must be submitted on a separate tape or set of cards and must be identified both on the tape and by an accompanying memorandum.

Documentation of each of the transaction types, describing the processing which the data goes through and indicating the limitations of each type of transaction has been provided to the Regional Office by MDAD (Slaymaker's memorandum of June 6, 1973).

The Regional Office should use the previously discussed procedures to determine if identified suspect data should be updated, corrected or deleted by means of these transactions.

3.2 Future Areas of Responsibility

Future areas of Regional Office responsibility with respect to air quality data include:

a. Quality Control

Quality control practices in the operation of air monitoring instruments, laboratory analysis and data handling procedures is of the utmost importance in producing valid air quality data. The Regional Offices should therefore encourage quality control programs at the State and local level. To aid the Regional Offices in this effort, the Quality Assurance and Environmental Monitoring Laboratory, NERC/RTP, has and is developing various manuals describing in detail, procedures to be followed during the course of sampling analysis and data handling for various pollutants.[9a,b,c,d]

The Control Programs Development Division has developed a general guideline for State and local quality control programs entitled "Quality Control Practices in Processing Air Pollution Samples."[5] This guideline should help the Regional Office establish a general quality control program at the State and local level.

b. Edit and Validation Checks

When MDAD develops the data validation programs and turns both the editor and data validation programs over to the Regional Offices, it is expected that the Regional Offices will assume the lead in initiating edit and validation checks on the incoming data. High quality data should then be transmitted to the National Aerometric Data Bank via upgraded remote access computer terminals.

4. CURRENT TECHNIQUES FOR SIP PROGRESS EVALUATION

It is difficult to develop comprehensive guidance on exactly how to determine whether a control strategy will need to be revised. While there may be a few situations where it is obvious that a plan revision is necessary, in general it will be a difficult task to determine that a plan is inadequate to attain the standards prior to the established attainment date. The problem is to determine whether AQCR's are progressing satisfactorily in relation to the emission limitations contained within the SIP. To this end, a Plan Revision Management System (PRMS) was developed to track the progress being made by States in implementing their SIP. PRMS provides a means for effectively combining information contained in SAROAD (air quality) NEDS (source emissions), and CDS (enforcement and compliance information) to compare measured progress against expected progress.

This system is designed to monitor the progress of actual air quality levels, obtained from the quarterly reports, in relation to the anticipated air quality reductions which should occur as a result of compliance with approved emission limitations. If the difference between the observed and projected air quality levels exceed certain specified limits, then the site is "flagged" as a "potential problem." A number of flagging levels or tolerance limits are incorporated in the system to indicate that the site either has acceptable progress or is having a minor, moderate, or major problem toward attainment of the NAAQS. The tolerence limits were developed through the application of statistical quality control techniques which allow for the many variables associated with measured air quality concentrations. (See Figure 9)

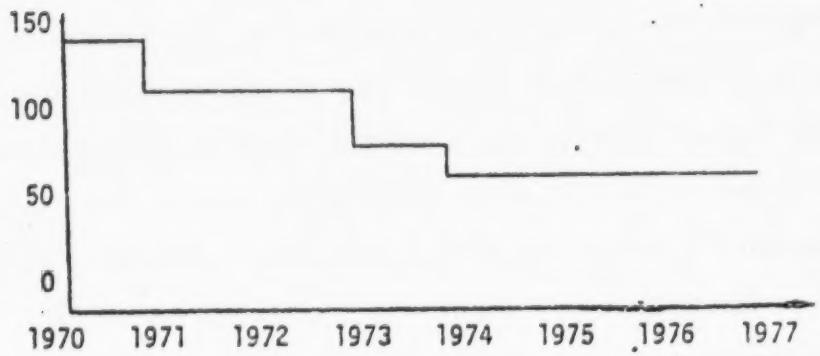
Figure 9

PLAN REVISION MANAGEMENT SYSTEM

Particulate Matter

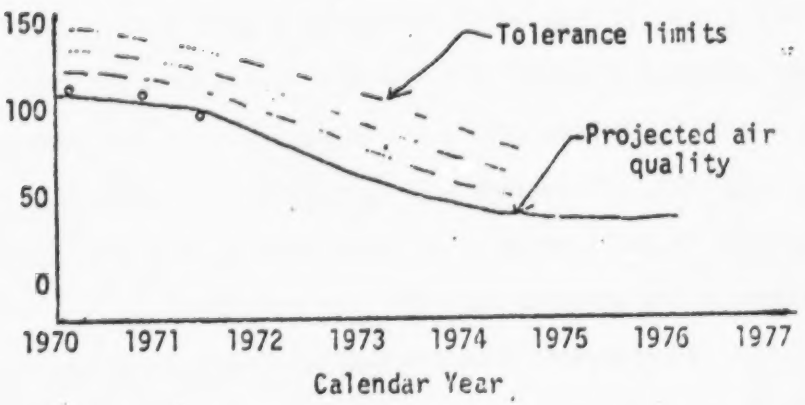
Emissions

(1000 tons/year)



Air quality

($\mu g/m^3$)



°Measured air quality

Step

#1 Calculation of emission reduction (NEDS, Emission Regulations)

#2 Review of compliance dates (SIP, CDS, Emission Regulations)

#3 Projection of air quality

#4 Establishment of tolerance limits or boundaries

#5 Measured air quality trend (SAROAD)

Reproduced from best available copy.

Once a "potential problem region" is identified, OAQPS will notify the appropriate Regional Office. This will be done on a semiannual basis. The Regional Office will be responsible for investigation and further assessment of the problem. The Regional Office should also report their findings to OAQPS indicating the action they have taken or plan to take.

While the PRMS will provide a mechanism to identify "potential problem regions" from an analytical point of view, the Regional Offices should be more intimately aware of the status of Regions within their States. Thus, the Regional Offices may be aware of other AQCR's not currently being analyzed by the PRMS which should be reviewed to determine if the plan is adequate to attain the NAAQS by the specified data for attainment.

Initially, there are 17 AQCR's contained in the PRMS. An additional 50 Regions were included in the system in January 1974. The additional 50 Regions that were selected for analysis were based on recommendations of the Regional Offices as to those AQCR's which should be reviewed to insure that adequate progress is being made toward attainment of the standards. By mid-1974, 50 more AQCR's are scheduled to be included in the PRMS. Thus, by July 1974, 117 Regions will be analyzed. The Regional Offices should indicate to OAQPS those AQCR's that they believe should be reviewed to determine the possible need for plan revisions.

It is understood that air quality levels throughout an AQCR are highly variable and that each monitoring site within the region must have levels at or below the national standards by the specified date for attainment to be in compliance with the Act. The PRMS analyzes all monitoring sites within SAROAD for the particular AQCR in question to

determine if adequate progress is being made. Thus, the system is capable of defining the problem on a much smaller scale than the entire AQCR. While most of the region may be showing adequate progress, a few sites, located in areas of maximum concentration, may be deviating from the desired air quality levels. Review of these sites will allow the Agency to take a much closer look at the real problem areas. Because the R.O. may only be required to review a very few problem sites, more effort can be placed upon those areas within an AQCR which appear to be having the most difficulty in attaining the standards. It is believed at this time that it will not be necessary in most cases to require a major plan revision for an entire AQCR. The revision or additional action can be tailored to a minimum number of sources to give the maximum amount of benefit toward attainment of the standards. Thus, a review to determine the adequacy of the progress for a region should be done on a site by site basis. The following two pages present the PRMS responsibilities and the associated action procedures.

ACTION PROCEDURES

A. Data Review Actions

1. The air quality data should be reviewed and work should proceed to certify the data if possible.
2. The monitoring site should be visited to determine if the monitor is properly located.
3. The meteorological conditions associated with the sampling period in questions should be reviewed to determine if any abnormal conditions could have effected the air quality levels.
4. The site location is source oriented and a unique projected curve for that site should be developed to better analyze the data.
5. A more detailed projected curve should be developed for the entire air quality control region.

B. Program Actions

1. A review of the compliance schedules for the AQCR should be conducted to determine if any sources have failed to meet any scheduled milestones or final compliance dates.
2. The State should be notified that a more effective implementation of the new source review procedures is needed to restrict growth in certain areas.
3. A special study should be initiated to determine the cause of the present air quality problem and the results are expected by _____.

C. Legal Actions

1. EPA/State enforcement action is necessary
2. Plan revisions is determined to be necessary and the State has been notified of the need for the revision.
3. The State's plan revision has not been submitted or approved and work has been initiated by EPA to develop the necessary

PRMS Responsibilities

OAQPS Responsibilities

- Calculate initial emission/time curve
- Develop initial projected air quality curve (Proportional model)
- Perform the computer analysis of measured vs projected air quality
- Notify each Regional Office of possible deficiencies
- Prepare a summary of the PRMS analysis for the Administrator's Progress Report
- Offer technical assistance to the Regional Office in investigating identified deficiencies
- If requested, rerun computer analysis with additional data provided by the Regional Office

Regional Office Responsibilities

- Investigate areas with possible deficiencies
- Inform OAQPS of the results of the investigation
- If a new projected air quality curve is determined to be necessary, it should be developed by the R.O.'s and submitted to OAQPS for a rerun of the PRMS analysis.
- If a plan revision is determined to be necessary by the R.O., inform the State of the type of revision necessary to correct the plan deficiency.

REFERENCES

1. SAROAD Users Manual, Office of Air Programs Publication No. APTD 0663, EPA, Research Triangle Park, N. C., July 1971.

2. SAROAD Terminal User's Manual, Office of Air Programs, Publication No. EPA-450/2-73-004, EPA, Research Triangle Park, N.C., October 1973.

3. "Field Operations Guide for Automatic Air Monitoring Equipment," Office of Air Programs, Publication No. APTD 0736, EPA, Research Triangle Park, N.C., November 1971.

4. "Guidelines for Technical Services of a State Air Pollution Samples," Office of Air Programs, Publication No. APTD 1347, EPA, Research Triangle Park, N.C., November 1972.

5. "Quality Control Practices in Processing Air Pollution Samples," Office of Air Programs, Publication No. APTD 1132, EPA, Research Triangle Park, N.C., March 1973.

6. Federal Register, Vol. 36, No. 228, November 25, 1971, page 22404.

7. Lowry, W.P. and R.W. Boubel, "Meteorological Concepts in Air Sanitation," Type-Ink., Corvallis, Oregon, 1967.

8. Symposium; Air Over Cities, Public Health Service, SEC Technical Report A-62-5, Cincinnati, Ohio, November 1961.

9. Guidelines for Development of a Quality Assurance Program, Office of Research and Monitoring, Quality Assurance and Environmental Monitoring Laboratory, Publication N.C. EPA-R4-73-028, EPA, Research Triangle Park, N.C., June 1973.

    a. Reference Method for the Continuous Monitors of Carbon Monoxide in the Atmosphere.

    b. Reference Method for the Determination of Suspended Particulates in the Atmosphere (High Value Method).

    c. Reference Method for Measurement of Photochemical Oxidants.

    d. Reference Method for the Determination of Sulfur Dioxide in the Atmosphere.

10. OAQPS #1.2-011 Guidelines for Determining the Need for Plan Revisions to the Control Strategy Portion of the Approved SIP.

11. Plan Revision Management System, System Summary, May 1974, USEPA, OAQPS, CPDD, Research Triangle Park, N.C.